Distribution Functions of Probabilistic Automata

Farrokh Vatan
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109
Farrokh.Vatan@jpl.nasa.gov

ABSTRACT

Each probabilistic automaton M over an alphabet $\mathcal{A}$ defines a probability measure $\mathtt{Prob}_M$ on the set of all finite and infinite words over $\mathcal{A}$. We can identify a k letter alphabet $\mathcal{A}$ with the set $\{0, 1, \ldots, k-1\}$, and, hence, we can consider every finite or infinite word w over $\mathcal{A}$ as a radix k expansion of a real number $X(w)$ in the interval $[0, 1]$. This makes $X(w)$ a random variable and the distribution function of M is defined as usual: $F(x) := \mathtt{Prob}_M\{\, w : X(w) < x \,\}$. Utilizing the fixed–point semantics (denotational semantics), extended to probabilistic computations, we investigate the distribution functions of probabilistic automata in detail. Automata with continuous distribution functions are characterized. By a new, and much more easier method, it is shown that the distribution function $F(x)$ is an analytic function if it is a polynomial. Finally, answering a question posed by D. Knuth and A. Yao, we show that a polynomial distribution function $F(x)$ on $[0, 1]$ can be generated by a probabilistic automaton iff all the roots of $F'(x) = 0$ in this interval, if any, are rational numbers. For this, we define two dynamical systems on the set of polynomial distributions and study attracting fixed points of random composition of these two systems.

Categories and Subject Descriptors

F.1.2 [**Computation by Abstract Devices**]: Models of Computation—*probabilistic computation*

General Terms

Theory, Verification

Keywords

probabilistic automata, denotational semantics, dynamical systems

1. INTRODUCTION

STOC '01 Crete, Greece

In this paper we use the fixed point semantics (denotational semantics) approach to probabilistic automata. In classical approach, properties of *states* are in the focus of attention; thus this approach leads to the theory of finite Markov chains (see, for example, [7, 10]). But, here we attend to external properties of automaton; we consider an automaton as a black box and look at states as far as they effect the external behavior of automaton. The classical theory of probabilistic automata is concerned with automata as *acceptors*, but we consider probabilistic automata as *generators* of random sequences. This approach in [6] is described as the following general problem: suppose that there is a source which provides uniform random bits, then by utilizing a computational device from a predetermined class of machines, which other distributions are possible to generate. In this paper, the authors consider several natural classes of machines, including the finite state automata.

Several other authors have considered similar problem. For example, in [5] the authors study the problem of generating the uniform distribution on the solution set of a given relation by probabilistic Turing machines which have access to unbiased coins. The problem studied in [4] can be described in the language of this paper as follows: generate finite probability distributions with finite computational trees where the probability of edges are restricted.

In this paper we present a detailed study of distribution function of probabilistic automata. Our approach is based on the fixed–point semantics. This method has been applied successfully in various areas of computer science (see, e.g., [3] and references there). We find the application of this method to probabilistic computation, based on the pioneering work [8], quiet natural and easy to apply. The main contribution of this paper is the complete solution of the problem posed in [6]: characterization of the polynomial distributions which can be generated by probabilistic automata. We prove that a polynomial distribution function $F(x)$ can be generated by a probabilistic automaton if and only if the only roots of $F'(x) = 0$ in the interval $[0, 1]$, if any, are rational numbers.

2. PROBABILITY MEASURES ON CPO'S

In this section we generalize the least fixed point approach (see, e.g., [9, 11] for details) to the nondeterministic computations. The nondeterministic computation can be considered without any probability, but considering probability distributions on the domain of states enable us to make into account various frequencies of possible outcomes and their statistical evaluations. First we give basic definitions and

results.

A partially ordered set (poset) $(D, \sqsubseteq)$ is said to be a **complete partially ordered set (cpo)** if it has a least element, denoted by $\perp$, and for every increasing sequence $(x_n)_{n\geq 1}$ of its elements the least upper bound (lub) of $(x_n)_{n\geq 1}$, denoted by $\sqcup_{n\geq 1} x_n$, does exist.

Let $(D, \sqsubseteq)$ and $(D', \sqsubseteq')$ be two cpo's. Let $f\colon D \longrightarrow D'$ be an increasing function; i.e., $f(x) \sqsubseteq' f(y)$ if $x \sqsubseteq y$. Then we say f is **continuous** if it preserve lub of increasing sequences; i.e., for every increasing sequence $(x_n)_{n\geq 1}$ it holds that $f(\sqcup_{n\geq 1} x_n) = \sqcup_{n\geq 1} f(x_n)$.

Let $(D, \sqsubseteq)$ be a cpo. A topology, called **Scott topology**, is defined on D in the following way. A subset $U \subseteq D$ is **open** if and only if whenever $x \in U$ and $x \sqsubseteq y$ then $y \in U$, moreover if the lub of an increasing sequence $(x_n)_{n\geq 1}$ belongs to U then $x_i \in U$ for some i [9, 11].

A **fixed point** of $f\colon D \longrightarrow D$, of a mapping of the cpo D to itself, is an element $x \in D$ such that $f(x) = x$. The **least fixed point** of f (if it exists) is the least element of the set of all fixed points of f. Let D be a cpo with the least element $\perp$ and $f\colon D \longrightarrow D$ be continuous. Then f has a *least fixed point* which actually is equal to $\sqcup_{n\geq 1} f^{[n]}(\perp)$; here $f^{[n]}$ is the n–fold iteration of f.

In this paper we consider a special cpo consisting of all finite and infinite words over a finite alphabet, but this restriction is not necessary; in [8] the general case of probability measure on an arbitrary cpo is studied.

Let $\mathcal{A}$ be a nonempty finite alphabet and $\mathcal{A}^*$ and $\mathcal{A}^\infty$ be the set of all finite and infinite words over $\mathcal{A}$, respectively. The empty word, denoted by $\perp$, is an element of $\mathcal{A}^*$. The **concatenation** of words x and y, denoted by xy, is defined if and only if x is a finite word. For every finite word x, we denote the length of x by $|x|$.

Let $D = \mathcal{A}^* \cup \mathcal{A}^\infty$ be ordered as follows: $x \sqsubseteq y$ if and only if x is a **prefix** of y; i.e., $y = xz$ for some $z \in D$. It is an easy fact that $(D, \sqsubseteq)$ is a cpo and the empty word is its least element. Note that if $x \sqsubseteq y$ and x is infinite then necessarily $x = y$.

Consider now the Scott topology on D, and let Ψ be the class of Borel sets in D; i.e., Ψ is the σ–algebra generated by open sets in D [1]. For every finite word $q \in \mathcal{A}^*$ let

$$V_q := \{\, x \in D : q \sqsubseteq x \,\}. \tag{1}$$

These elements of Ψ are our basic tool in studying the probabilistic automata. The subsets V_q, $q \in \mathcal{A}^*$, from a basis for the Scott topology on D.

A **probability measure** on D is a σ–additive function $P\colon \Psi \longrightarrow [0, 1]$ such that $P(D) = 1$. We denote the set of all probability measures on D by $\mathtt{Pr}(D)$.

For each probability measure P, we write "$P(x)$" instead of "$P(\{x\})$"; and we define the **support** of P as

$$\mathtt{supp}(P) := \{\, x \in D : P(x) > 0 \,\}.$$

If $P(\mathtt{supp}(P)) = 1$ then we represent P by

$$\{\, (x, P(x)) : x \in \mathtt{supp}(P) \,\}.$$

The probability measure P is called **finite measure** if $\mathtt{supp}(P)$ is a finite subset of $\mathcal{A}^*$ and $P(\mathtt{supp}(P)) = 1$. The set of all finite probability measures on D is denoted by $\mathtt{FPr}(D)$.

For probability measures P and P' on D we let

$$P \sqsubseteq_{\mathrm{P}} P'$$

if and only if for every open subset U of D we have $P(U) \leq P'(U)$.

In the sequel, we write simply "$P \sqsubseteq P'$" instead of "$P \sqsubseteq_{\mathrm{P}} P'$" whenever no ambiguity is possible.

In [8] it is proved that the ordered set $(\mathtt{Pr}(D), \sqsubseteq_{\mathrm{P}})$ is a cpo; its least element is $\{(\perp, 1)\}$, the probability measure concentrated on $\perp$.

THEOREM 2.1. *For P and P' in $\mathtt{Pr}(D)$, $P \sqsubseteq P'$ if and only if for every finite word $q \in \mathcal{A}^*$, $P(V_q) \leq P'(V_q)$. Moreover, if for every finite word $q \in \mathcal{A}^*$ we have $P(V_q) = P'(V_q)$ then $P = P'$.*

In [8] two important operations on $\mathtt{Pr}(D)$ are defined. Here we give their definitions and basic properties.

Definition 1. For $0 \leq p \leq 1$ and $P, P' \in \mathtt{Pr}(D)$, **the random selection between P and P' under the probability p** is the probability measure

$$R(p, P, P') := p \cdot P + (1 - p) \cdot P'.$$

We often denote this probability measure by "$p \cdot P + (1 - p) \cdot P'$".

Definition 2. Suppose that $f\colon D \longrightarrow D$ is a continuous function. The **probabilistic extension** of f, denoted by $\mathtt{ext}(f)$, is the function $\mathtt{ext}(f)\colon \mathtt{Pr}(D) \longrightarrow \mathtt{Pr}(D)$ defined as follows $\mathtt{ext}(f)(P)(B) := P(f^{-1}(B))$, for every $P \in \mathtt{Pr}(D)$ and $B \in \Psi$.

It is easy to show that $p \cdot P + (1-p) \cdot P'$ and $\mathtt{ext}(f)(P)$ are indeed probability measures. If f is a continuous function from D to itself, $p \in [0, 1]$ and P and P' are in $\mathtt{Pr}(D)$ then

$$\mathtt{ext}(f)\big(p \cdot P + (1 - p) \cdot P'\big) = p \cdot \mathtt{ext}(f)(P) + (1 - p) \cdot \mathtt{ext}(f)(P').$$

3. THE COMPUTATIONAL TREES

In [8] the author introduces a tree–wise approach to nondeterministic computations. Here we give an overview of his approach.

Consider a directed tree T. To each node v of T the output $e(v)$, which is a word over an alphabet $\mathcal{A}$, is corresponded. If the node v is not a leaf, the $e(v)$ should be a *finite* word. For r, the root of T, $e(r) = \perp$. Informally each node of T at depth n corresponds to a possible n^{th} step in the computation. For every edge (v, w) let a non–negative probability $p(v, w)$ be assigned such that whenever $w_1, \ldots, w_m$ are all immediate successors of a node v then

$$\sum_{i=1}^{m} p(v, w_i) = 1. \tag{2}$$

We call such tree a **computational tree**.

Let r be the root of a computational tree T. For every node v of T there exists exactly one path from r to v, say $(v_0, \ldots, v_m)$, with $v_0 = r$ and $v_m = v$. Then the **evaluation** of v is the finite word

$$E(v) = e(v_0) \cdots e(v_m). \tag{3}$$

(Note the order of e's, also note that $E(r) = \perp$.) It is obvious that if there exists a path from v to w, then $E(v) \sqsubseteq E(w)$, where "$\sqsubseteq$" is the prefix ordering on the words in $\mathcal{A}^*$.

Let L_n be the set of all leaves of depth *at most* n on a computational tree T. So $L_n \subseteq L_{n+1}$. Let V_n be the union of L_n and the set of nodes of depth n on T. We define recursively the *finite* probability measure π_n on V_n: $\pi_0(r) = 1$, where r is the root of T; $\pi_{n+1}(v) = \pi_n(v)$, if $v \in L_n$; $\pi_{n+1}(v) = \pi_n(w) \cdot p(w,v)$, if $v \in V_{n+1}$ and (w,v) is an edge with $w \in V_n$. So $\pi_n(v)$ is the probability to reach the node v from the root r by passing over at most n edges. It is easy to check that π_n is a probability measure on V_n.

Recall that the function E, defined by the equation (3), represents the "output" of the computational tree. Now that via probability measures π_n we have defined the probability to reach to each node, it is natural to extend this probabilities to the outputs of these nodes. Thus we define probability measures p_n on $\mathcal{A}^*$ which correspond to successive evaluations at times $n = 0, 1, 2, \ldots$:

$$p_n(w) := \sum_{v : E(v) = w} \pi_n(v), \qquad w \in \mathcal{A}^*. \tag{4}$$

Then p_n is in fact in $\mathtt{FPr}(D)$.

LEMMA 3.1. $p_n \sqsubseteq_{\mathrm{P}} p_{n+1}$.

Since $(\mathtt{Pr}(D), \sqsubseteq)$ is a cpo, the increasing sequence $(p_n)_{n \geq 0}$ has a limit, which justifies the next definition.

Definition 3. Let T be computational tree, and the probability measure p_n (on D) be defined as *(4)*. Then **the probability measure of** T, denoted by "$\mathtt{Prob}_T$," is the lub of the increasing sequence $(p_n)_{n \geq 0}$;

$$\mathtt{Prob}_T := \bigsqcup_{n \geq 0} p_n.$$

The finite probability measure p_n is called the n^{th} **cross section of** $\mathtt{Prob}_T$.

We now define an ordering on the trees to make them a cpo.

Definition 4. Let T be a computational tree. Suppose that T' is a obtained from T by deleting *all* descendants of some nodes of T. We say that T' is a **full subtree of** T. For computational trees T_1 and T_2 we let $T_1 \sqsubseteq_{\mathrm{T}} T_2$ (or simply $T_1 \sqsubseteq T_2$) if and only if T_1 is isomorphic with a full subtree of T_2. (Here the isomorphism of trees is in the usual sense of graph theory.)

THEOREM 3.2. *The set of computational trees with ordering* $\sqsubseteq_{\mathrm{T}}$ *is a cpo.*

Definition 5. The **cross–section at depth** m of a computational tree T, denoted by $\mathtt{depth}(T; m)$, is obtained by deleting all nodes on T with depth $> m$. If T is a finite tree of depth m, then we let $\mathtt{depth}(T; k) = T$, for $k \geq m$.

The following theorem shows a natural relation between the n^{th} cross section p_n of $\mathtt{Prob}_T$, defined in Definition 3, and the probability of cross section trees of T.

THEOREM 3.3. *Let* T *be a computational tree and* $T_n = \mathtt{depth}(T; n)$. *If* p_n *is the* n^{th} *cross–section of* $\mathtt{Prob}_T$ *then* $p_n = \mathtt{Prob}_{T_n}$.

THEOREM 3.4. *The function that corresponds each computational tree* T *to its probability measure* $\mathtt{Prob}_T$ *is continuous.*

Now we define an important operation on computational trees which enable us to construct new computational trees by "combining" the old ones.

Definition 6. Let $T_1, \ldots, T_n$ be computational trees with roots $r_1, \ldots, r_n$, respectively. Suppose that $w_1, \ldots, w_n$ are *finite* words over the alphabet $\mathcal{A}$, and $t_1, \ldots, t_n$ are non-negative numbers such that $t_1 + \cdots + t_n = 1$. Then the **direct sum** of T_i's with respect to w_i's and t_i's is the computational tree T with output function e, edge–probability function p and the root r such that $r_1, \ldots, r_n$ are the only immediate successors of r, $e(r_i) = w_i$, $p(r, r_i) = t_i$ and T contains all descendants of r_i in T_i with the same outputs and edge–probabilities. We denote T by

$$t_1 \Delta(w_1)(T_1) \oplus \cdots \oplus t_n \Delta(w_n)(T_n).$$

To study the probability measure of a direct sum, we use the following useful lemma and notation.

LEMMA 3.5. *Let* q *be a finite word over the alphabet* $\mathcal{A}$. *We remind that* $D = \mathcal{A}^* \cup \mathcal{A}^\infty$. *Then the function* $f_q \colon D \longrightarrow D$ *defined by* $f_q(x) = qx$ *is* continuous.

Definition 7. We denote the probabilistic extension of f_q by $\nabla(q)$. More explicitly,

$$\nabla(q) \colon \mathtt{Pr}(D) \longrightarrow \mathtt{Pr}(D)$$

such that for every probability measure P in $\mathtt{Pr}(D)$ and every open set U in D we have

$$\begin{aligned} \nabla(q)(P)(U) &:= P({f_q}^{-1}(U)) \\ &= P\{x \in D : qx \in U\}. \end{aligned}$$

In the important case when $U = V_z$, for some finite word $z \in \mathcal{A}^*$, we have

$$\nabla(q)(P)(V_z) = \begin{cases} 1 & \text{if } z \sqsubseteq q, \\ P(V_y) & \text{if } z = qy, \\ 0 & \text{otherwise.} \end{cases}$$

THEOREM 3.6. *Let* $0 < t < 1$, w *and* w' *be finite words over the alphabet* $\mathcal{A}$, *and*

$$S = t\Delta(w)(T) \oplus (1-t)\Delta(w')(T'),$$

where T *and* T' *are computational trees. Then*

$$\mathtt{Prob}_S = t\nabla(w)(\mathtt{Prob}_T) + (1-t)\nabla(w')(\mathtt{Prob}_{T'}).$$

The above theorem can easily be generalized to the direct sum of a finite set of computational trees.

THEOREM 3.7. (The Completeness Theorem) *For every probability measure* P *on* D *there exists a computational tree* T *such that* $P = \mathtt{Prob}_T$.

PROOF. For simplicity, let us assume that the domain D is based on the alphabet $\mathcal{A} = \{0, 1\}$; the proof can easily be generalized to an arbitrary alphabet. At depth 0, the only node is of course r, the root of T, with $e(r) = \bot$. If $P = \{(\bot, 1)\}$ then r is the only node of T; otherwise r is not a leaf.

Now let v be a node of T at depth n which is not a leaf, and $E(v) = q$, where q is a finite word over $\mathcal{A}$ and the output function E is defined by (3). Then the construction of T implies that $P(V_q) \neq 0$. We consider three temporary successors $v_\perp$, v_0 and v_1 for v with

$$e(v_\perp) = \perp, \qquad e(v_0) = 0, \qquad e(v_1) = 1,$$

$$\left.\begin{aligned} p(v, v_\perp) &= p_\perp = \tfrac{P(q)}{P(V_q)}, \\ p(v, v_0) &= p_0 = \tfrac{P(V_{q0})}{P(V_q)}, \\ p(v, v_1) &= p_1 = \tfrac{P(V_{q1})}{P(V_q)}. \end{aligned}\right\} \tag{5}$$

Whenever the probability of the edges (v, v_i), $i = \perp, 0, 1$, is zero then we delete that edge and the corresponding node v_i. The node $v_\perp$ is a leaf, and also if $p(v, v_\perp) = 1$, then we delete the edge $(v, v_\perp)$ and the node v will turn to be a leaf of T.

Since $P(V_q) = P(q) + P(V_{q0}) + P(V_{q1})$, hence the condition (2) is satisfied, and T is in fact a computational tree.

It is clear from the definition of T that if v is a node of T at depth $n+1$ then $|E(v)| = n$ or $n+1$. It is simple to verify that for every $n \geq 1$, if $T_n = \mathtt{depth}(T; n)$ then $\mathtt{Prob}_{T_n}(V_q) = P(V_q)$, whenever $|q| < n$. Consequently, $\mathtt{Prob}_T(V_q) = \sup_n \mathtt{Prob}_{T_n}(V_q) = P(V_q)$. Therefore, by Theorem 2.1, $\mathtt{Prob}_T = P$. □

4. THE PROBABILISTIC AUTOMATA

Definition 8. A **probabilistic automaton** is a quadruple

$$M = (Z, \mathcal{A}, S_1, \Pi) \tag{6}$$

where Z is a finite nonempty set of **states**, $\mathcal{A}$ is a finite nonempty **alphabet**, $S_1 \in Z$ is a specified state called the **initial state**, and Π is the probabilistic transition relation consisting of quadruples (S, S', q, t) where S and S' are states, $q \in \mathcal{A}^*$ is a finite word over the alphabet $\mathcal{A}$, and t is a real number such that $0 < t \leq 1$. Moreover, if (S, S'_i, q_i, t_i), $i = 1, \ldots, m$, are all members of Π such that their first component is S, then $t_1 + \cdots + t_m = 1$. We also assume that for every S, S' and q there exists at most one t such that $(S, S', q, t) \in \Pi$; moreover, we assume that Π does not contain any quadruple of the form $(S, S', \perp, 1)$.

For states S and S', we let $S \longrightarrow S'$ if and only if for some q and t the quadruple (S, S', q, t) is a member of Π. We say S' **is accessible from** S if $S' = S$ or there exists a sequence $W_1, \ldots, W_m$ of states such that $W_1 = S$, $W_m = S'$ and $W_i \longrightarrow W_{i+1}$, for $i = 1, \ldots, m-1$. *We assume that in every probabilistic automaton all states are accessible from the initial state.*

A state S of the probabilistic automaton (6) is called a **non–terminal state** if there are $S' \in Z$, $q \in \mathcal{A}^*$ and $0 < t \leq 1$ such that $(S, S', q, t) \in \Pi$. Otherwise, we call S a **terminal state**, i.e., if S is terminal then S is *not* the first component of any quadruple in Π.

There exists a natural connection between probabilistic automata and a subclass of computational trees. To show this connection, we first define the tree of an automaton.

Definition 9. Let $M = (Z, \mathcal{A}, S_1, \Pi)$ be a probabilistic automaton. The **computational tree of** M, denoted by $\mathtt{Tree}(M)$, is defined as follows. Each node v of $\mathtt{Tree}(M)$ has a label $\ell(v)$ which is a state in Z. The label of the root of $\mathtt{Tree}(M)$ is S_1, the initial state. If a node v of $\mathtt{Tree}(M)$ has label $\ell(v) = S \in Z$ and $(S, S'_1, q_1, t_1), \ldots, (S, S'_m, q_m, t_m)$ are all quadruples in Π such that $S \longrightarrow S'_i$, then v has exactly m immediate successor $v_1, \ldots, v_m$ with $\ell(v_i) = S'_i$, $e(v_i) = q_i$ and $p(v, v_i) = t_i$, for all $1 \leq i \leq m$.

It is easy to check that $\mathtt{Tree}(M)$ is actually a computational tree.

Definition 10. Let T be a computational tree and v be a node of T. In an obvious way, v and all its descendants form a computational tree with v as its root. We denote this tree by $T(v)$ and we call it **the subtree of** T **originated at** v. For nodes u and v of T, we say u and v are **equivalent** and write $u \cong v$ if and only if $T(u)$ and $T(v)$ are isomorphic; i.e., they differ at most by interchanging the order of edges.

Clearly $\cong$ is an equivalence relation on the set of nodes of T. The set of equivalence classes is simply denoted by $T/\cong$. We now have the following theorem which characterizes the computational trees associated with probabilistic automata.

THEOREM 4.1. *Let T be a computational tree. Then $T/\cong$ is a finite set if and only if there exists a probabilistic automaton M such that T and $\mathtt{Tree}(M)$ are isomorphic.*

Once we have defined the computational tree of a probabilistic automaton, we can extend the notions associated with computational trees to probabilistic automata.

Definition 11. Let M be a probabilistic automaton. Then **the probability measure of** M is $\mathtt{Prob}_{\mathtt{Tree}(M)}$; i.e., the probability measure generated by the computational tree of M. We denote this probability measure simply by $\mathtt{Prob}_M$.

We now investigate the probability measure of probabilistic automata in detail. Let $M = (Z, \mathcal{A}, S_1, \Pi)$ be a probabilistic automaton, and $Z = \{S_1, \ldots, S_z\}$. We can obtain various automata by considering various states of M as initial state. More specifically, let

$$M_i = (Z_i, \mathcal{A}, S_i, \Pi_i), \tag{7}$$

where Z_i is the set of states accessible from S_i, and Π_i is the restriction of Π to Z_i (hence $M_1 = M$). For every state S_i, $i = 1, \ldots, z$, suppose that all quadruples in Π whose first component is S_i are as follows:

$$\begin{gathered} \left(S_i, S_{g_i(1)}, q^i_1, t^i_1\right), \\ \vdots \\ \left(S_i, S_{g_i(z_i)}, q^i_{z_i}, t^i_{z_i}\right). \end{gathered}$$

If $P_i = \mathtt{Prob}_{M_i}$ then, by Theorem 3.6,

$$P_i = \sum_{j=1}^{z_i} t^i_j \nabla(q^i_j)\left(P_{g_i(j)}\right). \tag{8}$$

Note that if the state S_i is a *terminal state*, i.e., S_i is not the first component of any member of Π, then $\mathtt{Tree}(M_i)$ is the empty computational tree and $P_i = \{(\perp, 1)\}$, the probability measure concentrated on $\perp$.

We now look at (8) from another point of view. Consider the cpo D of all finite and infinite words over the alphabet $\mathcal{A}$.

Definition 12. The **continuous function of the probabilistic automaton** M with probabilities defined by (8) is the continuous function

$$\Phi_M : (\mathtt{Pr}(D))^z \longrightarrow (\mathtt{Pr}(D))^z$$

defined as follows

$$\Phi_M = (\varphi_1, \ldots, \varphi_z), \tag{9}$$

$$\varphi_i(Q_1, \ldots, Q_z) = \sum_{j=1}^{z_i} t_j^i \nabla(q_j^i)\left(Q_{g_i(j)}\right), \tag{10}$$

and if S_i is a terminal state then

$$\varphi_i(Q_1, \ldots, Q_z) = Q_i. \tag{11}$$

From (8) it follows that $(P_1, \ldots, P_z)$ is a fixed point of Φ_M; we show that it is in fact the *least fixed point* of Φ_M.

THEOREM 4.2. *Let $M = (Z, \mathcal{A}, S_1, \Pi)$ be a probabilistic automaton with $Z = \{S_1, \ldots, S_z\}$. Suppose that the probabilistic automaton M_i is defined as (7). Let P_i, defined by (8), be the probability measure of M_i, and let $\Phi_M = (\varphi_1, \ldots, \varphi_z)$, defined by (9)–(11), be the continuous function of M. Then $(P_1, \ldots, P_z)$ is the* least fixed point *of Φ_M.*

Remark. In general the equations (8) from a system which may be called **stochastic system of linear homogeneous equations** over $\mathtt{Pr}(D)$. Conversely, any such system can be associated to a probabilistic automaton and its least solution is the probability measure of that automaton.

5. THE NORMAL FORM OF PROBABILISTIC AUTOMATA

Before we start studying the properties of probabilistic automata, we define a normal form of these automata. This normal form will facilitate the future proofs. The probabilistic automata M and M', over the same alphabet $\mathcal{A}$, are called **equivalent** if for every finite word $q \in \mathcal{A}^*$,

$$\mathtt{Prob}_M(V_q) = \mathtt{Prob}_{M'}(V_q).$$

First we show that it is possible to get rid of probability–one and empty–output edges.

LEMMA 5.1. *Every probabilistic automaton M is equivalent to another automaton M' such that there is an edge of the form $(S, S', w, 1)$ in M' only if S is the initial state.*

LEMMA 5.2. *For every probabilistic automaton M there is an equivalent automaton M' such that M' has no edge with the empty output, except the edges of the form $(S_1, S, \perp, t)$ where S_1 is the initial state and S is a terminal state.*

6. PROBABILITY OF LARGE OUTPUTS

In this section we study some basic properties of probability measures of probabilistic automata.

THEOREM 6.1. *Let $\Phi_M = (\varphi_1, \ldots, \varphi_z)$ be the continuous function of a probabilistic automaton $M = (\Sigma, \mathcal{A}, S_1, \Pi)$ defined by (10) and $(P_1, \ldots, P_z)$ be its least fixed point. Suppose that for every $i = 1, \ldots, z$, if S_i is not a terminal state, there exists a finite word q_i such that $0 < P_i(V_{q_i}) < 1$. Then there exists a real number $0 < \tau < 1$ and a positive integer μ such that for every finite word q, every $i = 1, \ldots, z$ and every integer $k \geq 1$*

$$\text{if} \quad |q| \geq k \cdot \mu \quad \text{then} \quad P_i(V_q) \leq \tau^k. \tag{12}$$

Definition 13. A probability measure P is **degenerate** if $P(V_q)$ is either 0 or 1 for every finite word q. A probabilistic automaton is a **degenerate automaton** if $\mathtt{Prob}_M$ is degenerate.

In the sequel, all automata are non–degenerate, unless the contrary is explicitly stated.

THEOREM 6.2. *A probability measure P is degenerate if and only if P is a probability measure concentrated on a (finite or infinite) word w.*

7. DISTRIBUTION FUNCTIONS OF PROBABILISTIC AUTOMATA

We identify a k letter alphabet $\mathcal{A}$ with the set of digits $\{0, 1, \ldots, k-1\}$. Every (nonempty) finite or infinite word w over $\mathcal{A}$ can be considered as a radix k expansion of a real number $X_k(w)$ in the interval $[0, 1]$. More explicitly, if $w = a_1 a_2 \cdots$

$$X_k(w) := \sum_j \frac{a_j}{k^j}. \tag{13}$$

If k, the size of the alphabet, is clear, we simply write "X" instead of "X_k".

For every finite nonempty word q, we let $R(q)$ to be the set of all real numbers whose representation begins with q; i.e., if $q = a_1 a_2 \cdots a_m$ then

$$R(q) := \left[\frac{a_1 k^{m-1} + a_2 k^{m-2} + \cdots + a_m}{k^m},\ \frac{a_1 k^{m-1} + a_2 k^{m-2} + \cdots + a_m + 1}{k^m}\right). \tag{14}$$

Not that for every finite word q, $X(q0) = X(q)$ but $R(q0) \neq R(q)$.

Let, as before, D be the cpo of all finite and infinite words over the alphabet $\mathcal{A}$. Then X is a mapping from D into $\mathbb{R}$.

LEMMA 7.1. *The mapping X maps V_q to $R(q)$; i.e., for every finite nonempty word q,*

$$V_q = \{ w \in D : X(w) \in R(q) \}.$$

Definition 14. Let M be a probabilistic automaton. The probability measure $\mathtt{Prob}_M$ makes $X(w)$ a random variable and **the distribution function of** M, $F_M : \mathbb{R} \longrightarrow [0, 1]$, is defined as follows

$$F_M(x) := \mathtt{Prob}_M\{ w \in D : X(w) < x \}.$$

More generally, if T is a computational tree, we can define the **the distribution function of** T, $F_T : \mathbb{R} \longrightarrow [0, 1]$, by replacing $\mathtt{Prob}_M$ by $\mathtt{Prob}_T$ in the above identity.

(Note that $F_M(x) = 0$ for $x \leq 0$ and $F_M(x) = 1$ for $x \geq 1$.)

In order to characterize the automaton with continuous distribution functions, we need the following theorem.

THEOREM 7.2. *The distribution function of a non–degenerate probabilistic automaton is continuous if and only if it has no terminal state.*

8. ANALYTICITY OF THE DISTRIBUTION FUNCTION

The following theorem is proved in [6]. In the final version of this paper, we shall give an alternate proof of it in our denotational semantics approach.

THEOREM 8.1. *Let $F_M(x)$ be the distribution function of a probabilistic automaton M with no terminal states. If $F_M(x)$ is analytic on the interval $[c,d] \subseteq [0,1]$, with $c < d$, then $F_M(x)$ is a polynomial on $[c,d]$.*

9. EQUIVALENCE OF AUTOMATA OVER DIFFERENT ALPHABETS

The results of the next sections are about probabilistic automata over the alphabet $\{0,1\}$. We notice that we will get the same results by considering automata over alphabets with more than two letters.

In general, this is not true that an automaton M_1 over the alphabet $\mathcal{A}_1$ can be simulated by some automaton M_2 over another alphabet $\mathcal{A}_2$. This may happens when M_1 has terminal states. We will give examples of this phenomenon in the final version of this paper. We say alphabets $\mathcal{A}_1$ and $\mathcal{A}_2$ are **equivalent** if for every probabilistic automaton M_1 *with no terminal state* over the alphabet $\mathcal{A}_1$ there is an automaton M_2 over the alphabet $\mathcal{A}_2$ such that $F_{M_1} = F_{M_2}$, and vice versa.

THEOREM 9.1. *All finite alphabets are equivalent.*

10. THE 0–PART AND 1–PART OF FINITE–STATE DISTRIBUTION FUNCTIONS

From now on, we suppose that the alphabet of all automata is $\{0,1\}$, and D is the cpo of all finite and infinite words on $\{0,1\}$

Definition 15. A distribution function $F(x)$ on $[0,1]$ is called a **finite–state distribution (fsd)** if there exists a probabilistic automaton M such that $F(x) = F_M(x)$.

Consider the random variable X, defined by equation (13), and distribution function F_T of a computational tree T. For every word $w \in D$, $X(0w) = \frac{1}{2}X(w)$, and $X(1w) = \frac{1}{2} + \frac{1}{2}X(w)$. From these relations it follows that if the computation tree T_1 is the direct sum of T_2 and T_3 of the form

$$T_1 = t\Delta(0)(T_2) \oplus (1-t)\Delta(1)(T_3),$$

and if $F_i(x)$ is the distribution function of T_i, $i = 1,2,3$, then

$$F_1(x) = \begin{cases} tF_2(2x) & \text{if } 0 \le x < \frac{1}{2}, \\ t + (1-t)F_3(2x-1) & \text{if } \frac{1}{2} \le x \le 1. \end{cases} \tag{15}$$

Definition 16. Every distribution function $F(x)$ on $[0,1]$ defines a probability measure P_F on D in the natural way. For every finite word $q = a_1 \cdots a_n \in D$, let

$$P_F(V_q) = F\left(\frac{a_1}{2} + \cdots + \frac{a_n}{2^n} + \frac{1}{2^n}\right) - F\left(\frac{a_1}{2} + \cdots + \frac{a_n}{2^n}\right).$$

Then P_F extends to a probability measure on D. By Theorem *3.7* there exists a computational tree T_F such that $P_F = \mathtt{Prob}_{T_F}$. We call P_F the **probability of the distribution** F and T_F the **tree of the distribution** F.

By Theorem 3.7 (Page) there exist computational trees T, T_0 and T_1 such that $T = T_F$ and

$$T = t\Delta(0)(T_0) \oplus (1-t)\Delta(1)(T_1),$$

where $t = F(\frac{1}{2})$. Let $F^{(i)}(x)$ be the distribution function of T_i, for $= 0,1$.

Definition 17. The distribution functions

$$F^{(0)}(x) \quad \text{and} \quad F^{(1)}(x)$$

are called the **0–part** and **1–part** of $F(x)$, respectively.

With this notation we have

$$F(x) = \begin{cases} F(\frac{1}{2})F^{(0)}(2x) & \text{if } 0 \le x < \frac{1}{2} \\ F(\frac{1}{2}) + (1 - F(\frac{1}{2}))F^{(1)}(2x-1) & \text{if } \frac{1}{2} \le x \le 1 \end{cases} \tag{16}$$

$$F^{(0)}(x) = \frac{F(\frac{x}{2})}{F(\frac{1}{2})} \qquad 0 \le x \le 1 \tag{17}$$

$$F^{(1)}(x) = \frac{F(\frac{x+1}{2}) - F(\frac{1}{2})}{1 - F(\frac{1}{2})} \qquad 0 \le x \le 1 \tag{18}$$

Definition 18. The 0–parts of $F^{(0)}(x)$ and $F^{(1)}(x)$ are denoted by $F^{(00)}(x)$ and $F^{(10)}(x)$, and the 1–parts are denoted by $F^{(01)}(x)$ and $F^{(11)}(x)$, respectively. Generally, for every finite nonempty word w over $\{0,1\}$, the w–**part** of $F(x)$ is denoted by $F^{(w)}(x)$, and the 0–part and 1–part of $F^{(w)}(x)$ are denoted by $F^{(w0)}(x)$ and $F^{(w1)}(x)$, respectively.

THEOREM 10.1. *Suppose that $F_1(x), \ldots, F_n(x)$ are distribution functions on $[0,1]$. Suppose that there exists an integer $m \ge 1$ such that every distribution function $F_i^{(w)}(x)$, for every $w \in \{0,1\}^*$ of length m and every $1 \le i \le n$, belongs to the convex hull of $\{F_1(x), \ldots, F_n(x)\}$. Then every $F_i(x)$ is an fsd.*

10.1 Generating distributions $J_n(x) = x^n$

In Section 8 we mentioned that the only analytic distribution on $[0,1]$ generated by probabilistic automata are polynomials. In this section we investigate the polynomial distributions $J_n(x) = x^n$ generated by these automata.

THEOREM 10.2. *For every $n \ge 1$, the polynomial distribution function $J_n(x) = x^n$ is an fsd.*

PROOF. Equation (16) implies

$$J_n(x) = \frac{1}{2^n}J_n^{(0)}(2x) + \frac{2^n - 1}{2^n}J_n^{(1)}(2x-1). \tag{19}$$

Equation (17) implies that $J_n^{(0)}(x) = J_n(x) = x^n$. To complete the proof we show that $J_n^{(1)}(x)$ is a convex combination of $J_j(x)$, $1 \le j \le n$. Actually we have

$$J_n^{(1)}(x) = \frac{1}{2^n - 1}\sum_{j=0}^{n-1}\binom{n}{j}J_{n-j}(x).$$

Now the result follows from Theorem 10.1. □

11. GENERATING POLYNOMIAL DISTRIBUTIONS

In this section we deal with a problem posed by D. Knuth and A. Yao [6]: "which distribution functions can be generated by probabilistic automata?" we solve this problem completely as follows: every distribution function $F(x)$ on $[0,1]$ is an fsd, except in the case that $F'(x)$ has a root at an *irrational* point of $[0,1]$. First we generalize the basic notions of discrete dynamical systems [2].

Definition 19. Let G be a region in the Euclidean space $\mathbb{R}^n$. A **random (discrete) dynamical system** on G is a pair (f_0, f_1) of mappings from G into G. For every finite word q over $\{0,1\}$, the q**–iterate** of a point $\boldsymbol{x} \in G$, denoted by $\boldsymbol{x}^{(q)}$, is defined recursively as follows

$$\boldsymbol{x}^{(\perp)} := \boldsymbol{x},$$
$$\boldsymbol{x}^{(qi)} := f_i\left(\boldsymbol{x}^{(q)}\right), \qquad i = 0, 1.$$

For every infinite word w over $\{0,1\}$, the w**–orbit** of a point $\boldsymbol{x} \in G$ is the set

$$\boldsymbol{x}^{(w)} := \left\{ \boldsymbol{x}^{(q)} : q \text{ is finite and } q \sqsubseteq w \right\}.$$

For example, $\boldsymbol{x}^{(0010)} = f_0\left(f_1\left(f_0\left(f_0\left(\boldsymbol{x}\right)\right)\right)\right)$.

For a point $\boldsymbol{p} \in G$, **the basin of attraction of** $\boldsymbol{p}$, denoted by $B(\boldsymbol{p})$, is the set of all $\boldsymbol{x} \in G$ such that for every infinite word $w \in \{0,1\}^\infty$ the w–orbit $\boldsymbol{x}^{(w)}$ (as a sequence) tends to $\boldsymbol{p}$. We say $\boldsymbol{p}$ is an **attracting point** if $B(\boldsymbol{p})$ is a neighborhood of $\boldsymbol{p}$.

So, if $\boldsymbol{x} \in B(\boldsymbol{p})$ then for *every* infinite word $w \in \{0,1\}^\infty$ if we let q_n be the prefix of w of length n then $\lim_{n\to\infty} \boldsymbol{x}^{(q_n)} = \boldsymbol{p}$.

Let $F(x) = a_1x + a_2x^2 + \cdots + a_nx^n$ be a distribution function on $[0,1]$. Let $\eta(F) = (a_1, a_2, \ldots, a_n)$. We **identify** $F(x)$ with the point $\eta(F)$ in the Euclidean space $\mathbb{R}^n$. We define $\mathcal{D}_n$ as the subset of $\mathbb{R}^n$ consisting of $\eta(F)$, where $F(x) = a_1x+a_2x^2+\cdots+a_nx^n$ is a distribution function; i.e., $(a_1, a_2, \ldots, a_n) \in \mathcal{D}_n$ iff $\sum_{i=1}^n a_ix^i$ is a distribution function on $[0,1]$.

In the sequel we use the notation $F(x) \in \mathcal{D}_n$ for $\eta(F) \in \mathcal{D}_n$. So $F(x) \in \mathcal{D}_n$ means that $F(x)$ is a polynomial distribution function on $[0,1]$ with degree $\leq n$; and if $(F_k(x))_{k\geq 1}$ is a sequence of distribution functions then $\lim_{k\to\infty} F_k(x) = G(x)$ means that, in the Euclidean metric on $\mathbb{R}^n$, for the corresponding points we have $\lim_{k\to\infty} \eta(F_k) = \eta(G)$.

For distribution function $F(x) = a_1x + a_2x^2 + \cdots + a_nx^n$, consider the distributions $F^{(0)}(x)$ and $F^{(1)}(x)$, as defined by the equations (17) and (18). Suppose that

$$F^{(0)}(x) = a_1'x + a_2'x^2 + \cdots + a_n'x^n,$$
$$F^{(1)}(x) = a_1''x + a_2''x^2 + \cdots + a_n''x^n,$$

and define the following mappings on $\mathcal{D}_n$:

$$\delta_n^{(0)}(a_1, a_2, \ldots, a_n) := (a_1', a_2', \ldots, a_n'),$$
$$\delta_n^{(1)}(a_1, a_2, \ldots, a_n) := (a_1'', a_2'', \ldots, a_n'').$$

In this section we study the random dynamical system $\left(\delta_n^{(0)}, \delta_n^{(1)}\right)$ in detail. Note that $\mathcal{D}_m$ is a proper subset of $\mathcal{D}_n$ for $m < n$, and $\delta_m^{(i)}$ $(i = 0, 1)$ is the restriction of $\delta_n^{(i)}$ to $\mathcal{D}_m$.

LEMMA 11.1. *For every $n \geq 1$, the set $\mathcal{D}_n$ is a compact convex subset of $\mathbb{R}^n$.*

Let $\mathcal{D}_{n,k}^0$ be the set of the points $(a_1, a_2, \ldots, a_n) \in \mathcal{D}_n$ such that $a_1 = \cdots = a_{k-1} = 0$ and $a_k \neq 0$. Note that if $(a_1, \ldots, a_n) \in \mathcal{D}_{n,k}^0$, then $a_k > 0$.

Let $\mathcal{D}_{n,k}^1$ be the set of the distribution functions $F(x) \in \mathcal{D}_n$ which satisfy the following condition: $\frac{d^k}{dx^k}F(x)\Big|_{x=1} \neq 0$ and $\frac{d^j}{dx^j}F(x)\Big|_{x=1} = 0$ for ever $1 \leq j \leq k-1$.

THEOREM 11.2. *The mapping $\delta_n^{(0)}$ has just n fixed points on $\mathcal{D}_n$ which are $J_k(x) = x^k$, $k = 1, \ldots, n$. For $1 \leq k \leq n-1$, the distribution $J_k(x)$ is an attracting fixed point and $J_n(x)$ is a repelling point. For $1 \leq k \leq n-1$, under the dynamical system $\delta_n^{(0)}$, the point $J_k(x)$ attracts all points on $\mathcal{D}_{n,k}^0$.*

PROOF. From the relation $F(x) = F(\frac{1}{2})F^{(0)}(2x)$, for $0 \leq x < \frac{1}{2}$, it easily follows that

$$\delta_n^{(0)}(a_1, a_2, \ldots, a_n) = \frac{1}{2^n F(\frac{1}{2})}\left(2^{n-1}a_1, 2^{n-2}a_2, \ldots, a_n\right). \quad (20)$$

From (20) it easily follows that the distribution functions $J_k(x) = x^k$ are fixed points of $\delta_n^{(0)}$. On the other hand, if $F(x) = a_1x + a_2x^2 \cdots + a_nx^n$ is a *fixed point* of $\delta_n^{(0)}$ then, by (20), we have $2^i \cdot p \cdot a_i = a_i$, for $i = 1, \ldots, n$, where $p = F(\frac{1}{2})$. Hence, if $a_\ell \neq 0$ then $F(\frac{1}{2}) = 2^{-\ell}$. Therefore, if for some ℓ, $a_\ell \neq 0$ then for every $k \neq \ell$ we have $a_k = 0$. Consequently, we have the following lemma

LEMMA 11.3. *The only fixed points of the mapping $\delta_n^{(0)}$ are*

$$J_1(x) = x, \quad J_2(x) = x^2, \quad \cdots \quad, J_n(x) = x^n.$$

Now some useful notations.

Definition 20. We write

$$(a_1, a_2, \ldots, a_n) \leq (b_1, b_2, \ldots, b_n)$$

if and only if $a_i \leq b_i$ for every $i = 1, \ldots, n$.

Definition 21. For a point $\boldsymbol{p} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$, if $a_1 \neq 0$ then the **slope–vector of** $\boldsymbol{p}$ is the $(n-1)$–dimensional vector

$$\mathrm{SV}(\boldsymbol{p}) = \left(\frac{a_2}{a_1}, \ldots, \frac{a_n}{a_1}\right).$$

For the point $\boldsymbol{p} = (a_1, a_2, \ldots, a_n) \in \mathcal{D}_n$, with $a_1 \neq 0$, the slope–vector of $\delta_n^{(0)}(\boldsymbol{p})$ is as follows

$$\mathrm{SV}\left(\delta_n^{(0)}(\boldsymbol{p})\right) = \left(\frac{a_2}{2a_1}, \frac{a_3}{4a_1}, \ldots, \frac{a_n}{2^{n-1}a_1}\right).$$

Therefore, $\mathrm{SV}\left(\delta_n^{(0)}(\boldsymbol{p})\right) \leq \frac{1}{2}\mathrm{SV}(\boldsymbol{p})$. Hence

$$\lim_{k\to\infty} \mathrm{SV}\left(\left(\delta_n^{(0)}\right)^{[k]}(\boldsymbol{p})\right) = (0, 0, \ldots, 0),$$

where $\left(\delta_n^{(i)}\right)^{[k]}$ denotes the k–fold iteration of $\delta_n^{(i)}$. Consequently, $\left(\delta_n^{(0)}\right)^{[k]}(\boldsymbol{p})$ converges to a point with the zero slope–vector. But the point $(1,0,\ldots,0)$ (i.e., the distribution function $J_1(x)=x$) is the only point in $\mathcal{D}_n$ with the zero slope–vector. Therefore, $J_1(x)=x$ is the only attracting fixed point of the mapping $\delta_n^{(0)}$ on the set $\mathcal{D}_n \setminus \mathcal{D}_n^0$, where

$$\mathcal{D}_n^0 := \{\,(a_1,a_2,\ldots,a_n)\in\mathcal{D}_n : a_1=0\,\}.$$

Note that $\mathcal{D}_n^0$ is part of the boundary of $\mathcal{D}_n$, because for every distribution function $F(x)=a_1x+a_2x^2+\cdots+a_nx^n$ we have $F'(0)=a_1\geq 0$. By an argument similar to the above one, it can be shown that all distribution functions $F(x)\in\mathcal{D}_{n,k}^0$ tend to $J_k(x)=x^k$. □

THEOREM 11.4. *The distribution functions $G_k(x)=1-(1-x)^k$, for $k=1,\ldots,n$, are the only fixed points of the mapping $\delta_n^{(1)}$ on $\mathcal{D}_n$. Each point tends to one of $G_i(x)$, $1\leq i\leq n-1$; and $G_n(x)$ is a repelling point. More specifically, under the dynamical system $\delta_n^{(1)}$, every point in $\mathcal{D}_{n,k}^1$ tends to $G_k(x)$.*

PROOF. First we prove the following useful lemmas. We remind the definition of $F^{(w)}(x)$, the w–part of a distribution function $F(x)$. We first derive a series of simple lemmas. All of them can be proved by simple inductions.

LEMMA 11.5. *For every $F(x)\in\mathcal{D}_n$, every $a\in[0,1]$ and every integer $k\geq 1$ we have*

$$\frac{d}{dx}F^{(1^k)}(a)=\frac{F'\left(\frac{a+2^k-1}{2^k}\right)}{2^k\left(1-F\left(1-\frac{1}{2^k}\right)\right)},$$

where $F'(x)$ is the first derivative of $F(x)$.

LEMMA 11.6. *For every $F(x)\in\mathcal{D}_n$ we have*

$$1-F\left(1-\frac{1}{2^k}\right)=\sum_{j=1}^{n}(-1)^{j+1}\frac{1}{j!2^{j\cdot k}}\left[\frac{d^j}{dx^j}\,F(x)|_{x=1}\right].$$

LEMMA 11.7. *For every $F(x)=a_1x+\cdots+a_nx^n\in\mathcal{D}_n$ and $1\leq m\leq n$ we have*

$$\frac{d^m}{dx^m}F(x)\Big|_{x=1-\frac{1}{2^k}}=\sum_{j=0}^{n-m}(-1)^j\frac{1}{j!2^{j\cdot k}}\left[\frac{d^{j+m}}{dx^{j+m}}F(x)\Big|_{x=1}\right].$$

LEMMA 11.8. *For every $F(x)\in\mathcal{D}_n$ and every integers $1\leq m\leq n$ and $k\geq 1$ we have*

$$\frac{d^m}{dx^m}F^{(1^k)}(x)\Big|_{x=0}=\frac{\frac{d^m}{dx^m}F(x)\Big|_{x=1-\frac{1}{2^k}}}{2^{m\cdot k}\left(1-F\left(1-\frac{1}{2^k}\right)\right)}.$$

Now we are ready to characterize the attracting points of the dynamical system $\delta_n^{(1)}$ on $\mathcal{D}_n$.

A simple calculation shows that $G_k^{(1)}(x)=G_k(x)$, so $G_k(x)$ is a fixed point of $\delta_n^{(1)}$.

First, suppose that $F(x)\in\mathcal{D}_{n,1}^1$; i.e. $F'(1)\neq 0$. Then, since

$$\lim_{k\to\infty}2^k\left(1-F\left(1-\tfrac{1}{2^k}\right)\right)=F'(1),$$

from Lemma 11.5 it follows that

$$\lim_{k\to\infty}\frac{d}{dx}F^{(1^k)}(x)=1,\qquad\text{for every } x\in[0,1].$$

Therefore, $G_1(x)=x$ is the only attracting fixed point of the mapping $\delta_n^{(1)}$ on $\mathcal{D}_{n,1}^1$.

Next, suppose that $F'(1)=0$. Let

$$z_j:=\lim_{k\to\infty}\frac{d^j}{dx^j}F^{(1^k)}(x)\Big|_{x=0}\qquad j=1,2,\ldots,n.$$

If $F(x)\in\mathcal{D}_{n,2}^1$, i.e., $F''(1)\neq 0$, then from Lemma 11.8, by applying l'Hôpital's rule, it follows that

$$\begin{aligned}z_1&=\lim_{k\to\infty}\frac{F'\left(1-\frac{1}{2^k}\right)}{2^k\left(1-F\left(1-\frac{1}{2^k}\right)\right)}\\&=\lim_{h\to 0}\frac{hF'(1-h)}{1-F(1-h)}\\&=\lim_{h\to 0}\left(2-\frac{hF'''(1-h)}{F''(1-h)}\right).\\&=2.\end{aligned}$$

A similar calculation implies

$$z_1=2,\qquad z_2=-2,\qquad z_3=z_4=\cdots=z_n=0.$$

This shows that if $F(x)\in\mathcal{D}_{n,2}^1$, then

$$\lim_{k\to\infty}F^{(1^k)}(x)=2x-x^2=G_2(x).$$

Also if $F(x)\in\mathcal{D}_{n,3}^1$, for the corresponding parameters z_j we have

$$z_1=3,\ z_2=-6,\ z_3=6,\ z_4=\cdots=z_n=0,$$

which implies

$$\lim_{k\to\infty}F^{(1^k)}(x)=3x-3x^2+x^3=G_3(x).$$

In general, if $F(x)\in\mathcal{D}_{n,m}^1$, for $1\leq m\leq n-1$, then, for $1\leq j\leq m$,

$$z_j=\lim_{k\to\infty}\frac{d^j}{dx^j}F^{(1^k)}(x)\Big|_{x=0}=(-1)^{j+1}\frac{m!}{(m-j)!}=\frac{d^j}{dx^j}G_m(x)\Big|_{x=0},$$

and $z_j=0$ for $m<j\leq n$. Thus $\lim_{k\to\infty}F^{(1^k)}(x)=G_m(x)$. □

The random dynamical system $\left(\delta_n^{(0)},\delta_n^{(1)}\right)$

Consider the set $\{0,1\}^k$ of binary words of length k. For every $w=c_1c_2\cdots c_k$, we define the *order* of w as the integer $\langle w\rangle=\sum_{j=1}^k c_j2^{k-j}$. So $\langle w\rangle$ is the position of w in the

lexicographic ordering on $\{0,1\}^k$. For example, the ordering imposed on $\{0,1\}^3$ by $\langle\cdot\rangle$ is as follows

$$000,\ 001,\ 010,\ 011,\ 100,\ 101,\ 110,\ 111.$$

Note that

$$\langle w0\rangle = 2\langle w\rangle, \qquad \langle w1\rangle = 2\langle w\rangle + 1. \tag{21}$$

LEMMA 11.9. *Suppose that $F(x) \in \mathcal{D}_n$ and $w \in \{0,1\}^\ell$ with $j = \langle w\rangle$. Then*

$$F^{(w)}(x) = \frac{F\left(\frac{x}{2^\ell} + \frac{j}{2^\ell}\right) - F\left(\frac{j}{2^\ell}\right)}{F\left(\frac{j+1}{2^\ell}\right) - F\left(\frac{j}{2^\ell}\right)}. \tag{22}$$

THEOREM 11.10. *The distribution function $J_1(x) = x$ is the attracting point of the random dynamical system*

$$\left(\delta_n^{(0)}, \delta_n^{(1)}\right)$$

on $\mathrm{Int}(\mathcal{D}_n)$, where $\mathrm{Int}(\mathcal{D}_n)$ is the interior of $\mathcal{D}_n$.

PROOF. Let $F(x) \in \mathcal{D}_n$. Then $F(x) \in \mathrm{Int}(\mathcal{D}_n)$ if and only if $F'(\alpha) > 0$ for every $\alpha \in [0,1]$, especially $J_1(x) \in \mathrm{Int}(\mathcal{D}_n)$. From Lemma 11.9 it follows that if for every $\alpha \in [0,1]$ we have $F'(\alpha) > 0$, then

$$\frac{d}{dx}F^{(w)}(x)\Big|_{x=\alpha} > 0,$$

for every $\alpha \in [0,1]$ and every finite word $w \in \{0,1\}^*$. Hence, if $F(x) \in \mathrm{Int}(\mathcal{D}_n)$ then $F^{(w)}(x) \in \mathrm{Int}(\mathcal{D}_n)$ for every $w \in \{0,1\}^*$.

Let $\varepsilon > 0$. The intersection of the closed half–spaces

$$F\left(\frac{1}{2^m}\right) \le \varepsilon + \frac{1}{2^m},$$

for $m = 1, 2, \ldots, n+1$ and $F(x) \in \mathcal{D}_n$, defines a polyhedron $\mathcal{C}_\varepsilon$ in $\mathbb{R}^n$ which contains $J_1(x) = x$. In other words, $\mathcal{C}_\varepsilon$ is the closed subset of the points $(a_1, \ldots, a_n) \in \mathbb{R}^n$ satisfying the following inequalities:

$$\sum_{i=1}^{n} \frac{1}{2^{im}} a_i \le \varepsilon + \frac{1}{2^m}, \qquad m = 1, 2, \ldots, n+1.$$

Then $\mathcal{C}_\varepsilon \subseteq \mathcal{C}_{\varepsilon'}$ if $\varepsilon \le \varepsilon'$; also we have $\bigcap_{\varepsilon>0} \mathcal{C}_\varepsilon = \{J_1(x)\}$. So, every neighborhood of $J_1(x)$ (i.e., every neighborhood of the point $(1, 0, \ldots, 0)$) in $\mathbb{R}^n$ contains some $\mathcal{C}_\varepsilon$ for sufficiently small $\varepsilon > 0$.

Let $w \in \{0,1\}^k$ and $j = \langle w\rangle$. Then from Lemma 11.9 it follows that

$$F^{(w)}\left(\frac{1}{2^m}\right) = \frac{F\left(\frac{1}{2^{k+m}} + \frac{j}{2^k}\right) - F\left(\frac{j}{2^k}\right)}{F\left(\frac{j+1}{2^k}\right) - F\left(\frac{j}{2^k}\right)}$$

$$= \frac{1}{2^m} \cdot \frac{F'(\alpha)}{F'(\beta)},$$

where

$$\frac{j}{2^k} \le \alpha \le \frac{1}{2^{k+m}} + \frac{j}{2^k},$$

$$\frac{j}{2^k} \le \beta \le \frac{j+1}{2^k}.$$

Thus, $0 \le |\alpha - \beta| \le \frac{1}{2^k}$. Since $F'(t) \ne 0$ for every $t \in [0,1]$, therefore,

$$\lim_{|w|\to\infty} F^{(w)}\left(\tfrac{1}{2^m}\right) = \tfrac{1}{2^m}.$$

Hence, for every finite word w with sufficiently large length, we have $F^{(w)}(x) \in \mathcal{C}_\varepsilon$, for small $\varepsilon > 0$. □

From equation (17) it follows that $F^{(0)}(x)$ can be obtained as follows: first restrict $F(x)$ to the closed interval $\left[0, \frac{1}{2}\right]$; we obtain an increasing function

$$\left[0, \tfrac{1}{2}\right] \longrightarrow \left[0, F\left(\tfrac{1}{2}\right)\right];$$

if we re–scale both axes such that the interval $\left[0, \frac{1}{2}\right]$ on the x–axis became $[0,1]$ and the interval $\left[0, F(\frac{1}{2})\right]$ on the y–axis became $[0,1]$, then we obtain the graph of $F^{(0)}(x)$. A similar construction, with restriction of $F(x)$ to $\left[\frac{1}{2}, 1\right]$ (accompanied by a re–scaling), leads to the graph of $F^{(1)}(x)$. By an inductive construction, the graph of $F^{(w)}(x)$, for binary word w, is obtained.

THEOREM 11.11. *If $H_1(x), \ldots, H_m(x)$ are fsd's, then any convex combination of them is also an fsd.*

LEMMA 11.12. *Each distribution function in $\mathrm{Int}(\mathcal{D}_n)$ is an fsd.*

PROOF. Let $H_1(x), \ldots, H_{n+1}(x)$ be vertices of a simplex $\mathcal{S}$ inside $\mathrm{Int}(\mathcal{D}_n)$ which contains $J_1(x) = x$. If $w \in \{0,1\}^*$ has sufficiently large length then, by Theorem 11.10, the distribution function $H_i^{(w)}(x)$, for every $1 \le i \le n+1$, belongs to the convex hull of $\{H_1(x), \ldots, H_{n+1}(x)\}$. Therefore, Theorem 10.1 implies that each $H_i(x)$ is an fsd.

For every $F(x) \in \mathrm{Int}(\mathcal{D}_n)$, Theorem 11.10 implies that $F^{(w)}(x)$ is in $\mathcal{S}$ for finite binary word w with sufficiently large length; and Theorem 11.11 implies $F^{(w)}(x)$ is an fsd. Thus, from Theorem 10.1 it follows that $F(x)$ is an fsd. □

We now study the distribution functions on the boundary of $\mathcal{D}_n$. We divide these functions to two classes: one is the class of functions $F(x)$ such that all roots of $F'(x)$ in the interval $[0,1]$ are *rational*, and the other class consists of the rest. We will show that all distribution functions in the first class are fsd, and the functions in the second class are not. For the proof in the case of each class we need new notions which we will develop.

LEMMA 11.13. *Let $F(x) \in \mathcal{D}_n$ and $F'(0) = 0$ or $F'(1) = 0$. Then $F(x)$ is an fsd.*

PROOF. We show that $F(x)$ can be generated by an automaton over the binary alphabet $\{0,1\}$. Fist suppose that $F'(0) = 0$. Then $F^{(w)}(x)$, for every $w = 0^k1$ with $k \ge 0$, is in $\mathrm{Int}(\mathcal{D}_n)$, and by Lemma 11.12 is an fsd. As discussed in the proof of Theorem 11.2, $F(x)$ tends to one of the fixed points $J_2(x) = x^2, \ldots, J_{n-1}(x) = x^{n-1}$, under the dynamical system $\delta_n^{(0)}$. The points $J_2(x), \ldots, J_{n-1}(x)$ are interior points of the intersection of $\mathcal{D}_n$ and the hyper–plane $a_1 = 0$. So, there are points $H_1(x), \ldots, H_t(x)$ in this intersection such that $J_2(x), \ldots, J_{n-1}(x)$ are in $\mathcal{C}$, the convex hull of $H_1(x), \ldots, H_t(x)$ and consequently $F(x)$ is an fsd.

A similar argument shows that all rational polynomial distribution functions $F(x)$ such that $F'(1) = 0$ are fsd. □

THEOREM 11.14. *If the distribution function $F(x)$ is on the boundary of $\mathcal{D}_n$ and every root of $F'(x)$ in the interval $[0,1]$ is* rational, *then $F(x)$ is an fsd.*

Definition 22. Let $F(x)$ be an fsd generated by the automaton $M = (\Sigma, \{0,1\}, S_1, \Pi)$ with $|\Sigma| = k$. For each $S_j \in \Sigma$, the automaton M_j is defined as the sub–automaton of M with S_j as its initial state. So $M_1 = M$. Let $F_j(x)$ be the distribution function of M_j. We call the set $\{F_1(x), \ldots, F_k(x)\}$ as a **set of the companions** of $F(x)$.

Definition 23. Let $F(x) \in \mathcal{D}_n$. We define $\mathcal{R}(F(x))$ to be the set of $\alpha \in [0,1]$ such that $\frac{d}{dx}F^{(w)}(x)\Big|_{x=\alpha} = 0$, for some $w \in \{0,1\}^*$.

LEMMA 11.15. *If $F(x) \in \mathcal{D}_n$, then $\mathcal{R}(F(x))$ is a finite set.*

THEOREM 11.16. *If $F(x) \in \mathcal{D}_n$ and $F'(\alpha) = 0$, for some* irrational *$0 < \alpha < 1$. Then $F(x)$ is* not *an fsd.*

PROOF. Let $\mathcal{B}_t$, $0 \le t \le 1$, be the set of all distribution functions $F(x) \in \mathcal{D}_n$ such that t is the only root of $F'(x)$ in the closed interval $[0,1]$. From equations (17) and (18) it follows that if $F(x) \in \mathcal{B}_t$ then the followings hold.

(i) if $0 \le t < \frac{1}{2}$ then $F^{(0)}(x) \in \mathcal{B}_{2t}$ and $F^{(1)}(x) \in \mathrm{Int}(\mathcal{D}_n)$;

(ii) if $\frac{1}{2} < t \le 1$ then $F^{(1)}(x) \in \mathcal{B}_{2t-1}$ and $F^{(0)}(x) \in \mathrm{Int}(\mathcal{D}_n)$.

Finally, if $t = \frac{1}{2}$ then $F^{(0)}(x) \in D^1_{n,1}$ and $F^{(1)}(x) \in D^0_{n,1}$ and in this case $F(x)$ is an fsd.

Consider now the following dynamical system on $[0,1]$;

$$b(x) = \begin{cases} 2x & \text{if } 0 \le x \le \frac{1}{2}, \\ 2x-1 & \text{if } \frac{1}{2} < x \le 1. \end{cases}$$

This is the well–known **baker map** and is *chaotic* on $[0,1]$ (see [2] for details); i.e., it has sensitive dependence on the initial condition, it has dense orbits, and its periodic points are dense in the interval $[0,1]$. It can be shown that all rational points in $[0,1]$ are periodic and every irrational point has dense orbit.

Let $t \ne \frac{1}{2}$ and $F(x) \in \mathcal{B}_t$. Then $F^{(0)}(x)$ or $F^{(1)}(x)$ is in $\mathcal{B}_{b(t)}$ if $0 \le t < \frac{1}{2}$ or $\frac{1}{2} < t \le 1$, respectively. Now, for every infinite word w, the w–path of $F(x)$ under the random dynamical system $\left(\delta_n^{(0)}, \delta_n^{(1)}\right)$ is in $\mathrm{Int}(\mathcal{D}_n)$ except for exactly one infinite word w_∞ which corresponds to the behavior of $b(t)$. More explicitly, if we let

$$\varepsilon_m = \begin{cases} 0 & \text{if } 0 \le b^{[m]}(t) \le \frac{1}{2}, \\ 1 & \text{if } \frac{1}{2} < b^{[m]}(t) \le 1, \end{cases}$$

then $w_\infty = \varepsilon_0\varepsilon_1\varepsilon_2\cdots$. Therefore,

$$\left\{\alpha \in [0,1] : \tfrac{d}{dx}F^{(\varepsilon_0\varepsilon_1\cdots\varepsilon_m)}(x)\Big|_{x=\alpha} = 0\right\} = \left\{b^{[m]}(t) : m = 0,1,2,\ldots\right\}.$$

If t is irrational then the set on the right–hand side is infinite, and Lemma 11.15 implies that $F(x)$ is *not* an fsd. □

Acknowledgments

This work was permformed in part at Electrical Engineering Department of UCLA and was supported by the Defense Advanced Research Projects Agency (DARPA) project MDA972–99–1–0017 [note that the content of this paper does not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred], and in part by the U.S. Army Research Office/DARPA under contract/grant number DAAD19–00–1–0172.

The research described in this paper also was performed partly at the Jet Propulsion Laboratory (JPL), California Institute of Technology, under contract with National Aeronautics and Space Administration (NASA). The Revolutionary computing Technologies Program of the JPL's Center for Integrated Space Microsystems (CISM) supported this work.

12. REFERENCES

[1] P. Billingsley. *Probability and Measure.* John Weily & Sons, New York, 3rd edition, 1995.

[2] R. E. Devany. *An Introduction to Chaotic Dynamical Systems.* the Benjamin/Cummings, New York, 1986.

[3] A. Edalat. Domain theory in stochastic processes. In *Proceedings of Tenth Annual IEEE Symposium on Logic in Computer Science*, pages 244–54, 1995.

[4] D. Feldman, R. Impagliazzo, M. Naor, N. Nisan, S. Rudich, and A. Shamir. On dice and coins: models of computation for random generation. *Information and Computation*, 104:159–174, 1993.

[5] M. R. Jerrum, L. G. Valiant, and V. V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, 1986.

[6] D. E. Knuth and A. C. Yao. The complexity of nonuniform random number generating. In J. F. Traub, editor, *Algorithms and Complexity: New directions and Recent Results*, pages 357–428, New York, 1976. Academic Press.

[7] A. Paz. *Introduction to Probabilistic Automata.* Academic Press, New York, 1971.

[8] N. Saheb-Djahromi. Cpo's of measure of nondeterminism. *Theoretical Computer Science*, 12:19–37, 1980.

[9] D. Scott. Outline of a mathematical theory of computation. In *Proceedings of 4th Annual Princeton Conference on Information Science and Systems*, pages 169–176, 1970.

[10] H. Starke. *Abstract Automata.* North–Holland, Amsterdam, 1972.

[11] J. E. Stoy. *Denotational Semantics: the Scott–Strachy Approach to Programming Language Theory.* MIT Press, Cambridge, MA, 1977.